ELSEVIER

Biochimica et Biophysica Acta 1453 (1999) 412–416

BIOCHIMICA ET BIOPHYSICA ACTA

BBA

Short sequence-paper

Cloning of the SmSPO-1 gene preferentially expressed in sporocyst during the life cycle of the parasitic helminth *Schistosoma mansoni*[1]

Daniela Ram, Frida Lantner, Etty Ziv, Vinca Lardans, Israel Schechter *

Department of Immunology, The Weizmann Institute of Science, Rehovot 76100, Israel

Received 8 December 1998; accepted 3 February 1999

Abstract

Schistosomes are parasitic helminths with a complex life cycle in human and snail hosts. They express stage-specific genes that conceivably determine distinct properties of the parasite at different developmental stages. Here we report the stage-specific gene SmSPO-1, which is preferentially expressed in sporocysts residing in the snail host. The cDNA and the gene were cloned and sequenced. The cDNA, from cap site to the poly(A) addition site, is 498 bp long. It encodes a protein of 117 amino acids with a hydrophobic signal peptide of 18 residues, indicating that SmSPO-1 is a secreted or a membranal protein. In the gene the cDNA is split into four exons spread over 2.1 kb of chromosomal DNA.

Keywords: Stage-specific gene; cDNA sequence; Gene sequence; Parasite life cycle; Sporocyst; (*Schistosoma mansoni*)

Bilharzia is a major health problem in large portions of Asia, Africa and South America. The disease is caused by parasitic helminths of the genus *Schistosoma* that have a complex life cycle involving human and snail hosts, as well as two larval stages living freely in water [1]. The transformation from one developmental stage to another is associated with major changes in size (0.1 to 20 mm), shape, physiological and biochemical properties of the parasite. Distinct properties of the parasite at different developmental stages are determined to a large extent by the activation/inactivation of a selected group of genes, termed stage-specific genes. Little is known on the regulation and function of stage-specific genes. To address these issues several genes expressed at different developmental stages of schistosome were cloned and characterized (for review see [2,3]), but to our knowledge a sporocyst-specific gene has not yet been reported. Here we report the cloning of the SmSPO-1 gene expressed preferentially in sporocysts.

The SmSPO-1 was identified by analysis of RNA from different developmental stages of the parasite: miracidium (living in water, infective to snail), sporocyst (parasitic form in snail), cercaria (emerging from snail, living in water, infective to human), schistosomulum (cercaria after penetration to human skin is transformed into schistosomulum), and adult worm (long-term parasitic form in human). The sporocysts reside in the hepatopancreas (HP) of the snail where they multiply asexually and metamorphose to cercariae. In infected HP about 50% of the RNA is parasite-specific [3]. Accordingly RNA

* Corresponding author. Fax: +972 (8) 934-4141; E-mail: lischech@weizmann.weizmann.ac.il

[1] Nucleotide sequences of SmSPO-1 reported in this paper have been submitted to GenBank under accession numbers AF109180 for the cDNA and AF109181 for the gene.

0925-4439/99/$ – see front matter
PII: S0925-4439(99)00012-5

from HP of infected snail served as the source of sporocyst RNA. RNA from the HP of uninfected snails and of mouse kidney served as controls.

The RNAs from different developmental stages of *Schistosoma mansoni* were translated in the wheat germ cell-free system and the protein products were reacted with rabbit antiserum to sonicate of cercarial tegument (SCT). The immunoprecipitates showed a strong protein band with an apparent molecular mass of 16 kDa in sporocyst but not in cercaria or adult worm, and a band of 22 kDa in all developmental stages tested; the control of HP of uninfected snail was free of any protein bands (Fig. 1A). These findings demonstrate that the mRNA encoding the 16 kDa protein, termed SmSPO-1, is stage-specific and is preferentially expressed in sporocyst.

The SmSPO-1 cDNA was cloned from a λgt11 cDNA library of sporocysts screened with rabbit antiserum to SCT. *Eco*RI digest of the clone released a 0.5 kb DNA fragment that was subcloned into pBR322 to yield pSmSPO-1. mRNA hybrid selection-immunoprecipitation experiments [4] showed specific hybridization of pSmSPO-1 with mRNA directing the synthesis of a 16 kDa protein. This mRNA was detected only in sporocyst but not in cercaria, adult worm and HP of uninfected snail. mRNA encoding the 22 kDa protein (Fig. 1A) was not detected by hybrid selection with pSmSPO-1 (Fig. 1B). The specificity of hybrid selection was further evidenced by hybridization of sporocyst RNA with pBR322 DNA that failed to select any mRNA translatable into a 16 kDa protein (Fig. 1B, lane 5). These findings establish that the SmSPO-1 cDNA clone encodes a sporocyst-specific protein with an apparent M_r of 16 000.

The intensity of the immunoprecipitated 16 kDa protein band (Fig. 1B) suggested that it was a major protein component. This was supported by the observation that protein products prior to immune precipitation showed a distinct 16 kDa band (Fig. 1C,

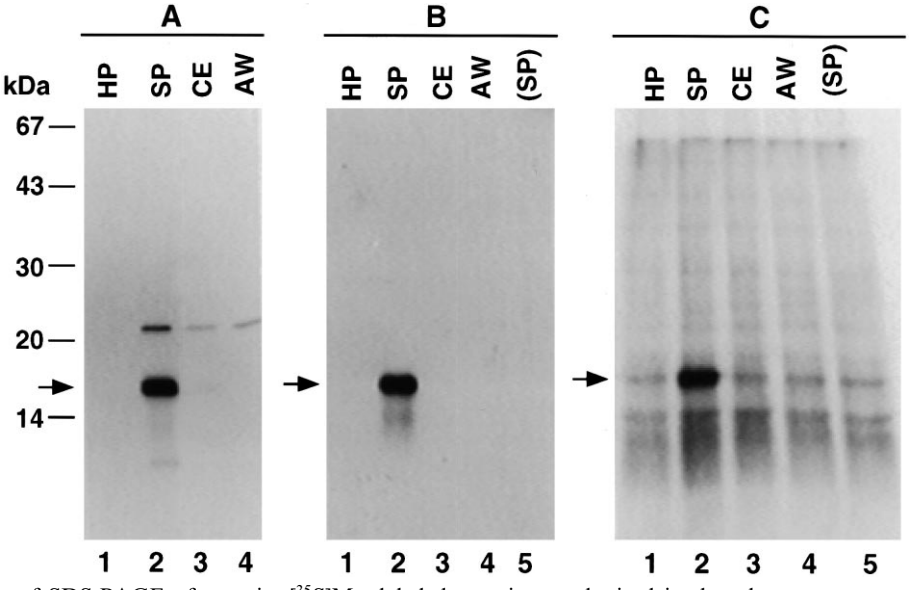

Fig. 1. Autoradiogram of SDS-PAGE of parasite [^{35}S]Met-labeled proteins synthesized in the wheat germ extract. (A) Cell-free products programmed by RNA of (1) hepatopancreas of normal snails (HP), (2) sporocysts (SP), (3) cercariae (CE) and (4) adult worms (AW) were immunoprecipitated by antiserum to SCT, and subjected to SDS-PAGE. (B) mRNA hybrid selection-immunoprecipitation. Total RNA of (1) HP of normal snails, (2) sporocysts, (3) cercariae and (4) adult worms was hybridized with pSmSPO-1 DNA immobilized on filter papers, annealed RNA was eluted, translated in the wheat germ extract, the cell-free products were reacted with antiserum to SCT, and immunoprecipitates were subjected to SDS-PAGE. Lane 5 (SP) represents sporocyst RNA treated as above except that hybridization was with immobilized pBR322 DNA. (C) mRNA hybrid selection. RNAs from different developmental stages were hybridized with immobilized pSmSPO-1 (lanes 1–4) or pBR322 (lane 5), elution of annealed RNA and translation of eluted RNA were done as in B, but the cell-free products were applied to SDS-PAGE without immunoprecipitation (B is autoradiogram after immunoprecipitation of proteins shown in C). The positions of protein size markers are indicated on the left. Arrow indicates the position of the 16 kDa SmSPO-1 protein.

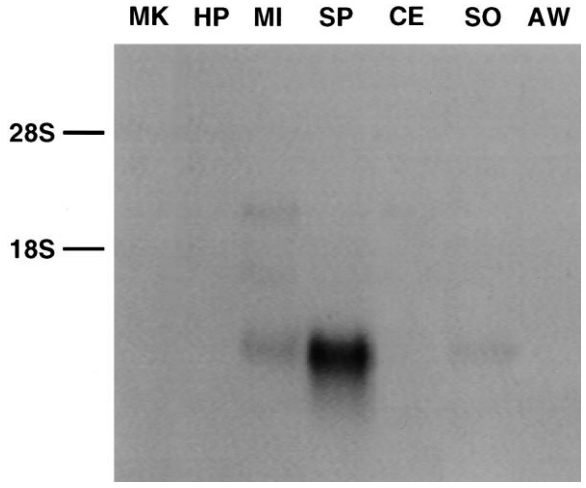

Fig. 2. Stage-specific expression of the SmSPO-1 mRNA. Northern blot (10 μg total RNA per lane) probed with the ^{32}P-labeled SmSPO-1 cDNA. RNA was prepared from mouse kidney (MK), hepatopancreas of normal snails (HP), miracidia (MI), sporocysts (SP), cercariae (CE), schistosomula (SO) 2 h after transformation from cercariae, and adult worms (AW). Positions of 28S and 18S rRNA are indicated.

lane 2). We mention that hybrid selection experiments with other proteins (paramyosin [2] and hsp70 [3] of schistosome, Ig light chains in the spleen [5], and others) revealed the appropriate protein only after immune precipitation (as in Fig. 1B) but not before immune precipitation (as in Fig. 1C).

Northern blots showed high levels of the 0.7 kb SmSPO-1 mRNA in sporocysts, small amounts in miracidia and smaller amounts in schistosomulum. The mRNA was undetectable in adult worm and cercaria (Fig. 2). Overexposure revealed trace amounts of the 0.7 kb SmSPO-1 mRNA in cercaria but not in adult worm. Two larger bands of approx. 1.7 and 2.8 kb detected in miracidia presumably represent unprocessed pre-mRNA (Fig. 2). By overexposure these bands were also observed in sporocyst, cercaria and schistosomulum. Controls of uninfected HP and mouse kidney did not show any hybridizing band with the SmSPO-1 cDNA probe (Fig. 2).

The SmSPO-1 cDNA (0.5 kb) was subcloned into the KS Bluescript plasmid and sequenced using the chain termination method [6]. The sequence (494 bp) showed an open reading frame of 351 bases starting with ATG (positions 1–3) encoding the first Met residue and ending with TCT (positions 349–351) encoding the C-terminal Ser residue, followed by a TAG termination codon. The 5′ end of the cDNA contained 12 bases upstream of the first ATG (Fig. 3). To determine the 5′ end of the mRNA a 22-mer oligonucleotide (complementary to position 25–46, Fig. 3) was annealed with sporocyst RNA for primer extension sequencing analysis [4]. The results showed 5′ extension by 17 nucleotides beyond the 5′ end of the cDNA, terminating with an A at position −29 (Fig. 3). The extended sequence was identical to the sequence of the gene in this region. Eukaryotic mRNA transcription is usually initiated by a purine, mostly A, surrounded by two or more pyrimidines [7]. The sequence around the A at position −29 is TTTAT (see Fig. 3 and gene sequence) in complete agreement with the structure of a cap site. Furthermore, the gene sequence revealed that this A is 26 bases downstream of a TATA box, in agreement with the finding that most cap sites are situated 21–27 bp from the TATA box [7]. It is likely that the first ATG is the initiator Met codon because it is the first ATG from the cap site, and it is preceded by two in-frame TGA termination codons (Fig. 3). The SmSPO-1 mRNA (0.7 kb, Fig. 2) is bigger than the cDNA (494 bp, Fig. 3), probably because the cDNA contains a short poly-A tail of only 17 adenines (Fig. 3) that in schistosome may extend to about 150 adenines [4], and 17 bp are missing from the 5′ untranslated region of the cDNA.

The open reading frame encodes a protein of 117 amino acids with a calculated molecular mass of 13 651 Da which is smaller than the size of 16 kDa estimated by gel electrophoresis (Fig. 1). This discrepancy is not unexpected since deviations of migration in SDS gels were observed for small proteins [8]. The predicted amino acid sequence revealed a hydrophobic signal peptide evidenced by visual inspection and by using the Kyte and Doolittle hydrophobicity plot [9]. Accordingly the cDNA sequence encodes a precursor that retains the signal peptide in a cell-free system [10], but in vivo the signal peptide is split to yield a mature protein. The SignalP program for signal peptides [11] indicates cleavage between Thr^{18} and Leu^{19}, to yield a mature protein of 99 residues (amino acids 19–117, Fig. 3) with a calculated molecular mass of 11 741 Da and p*I* 8.9. The presence of a signal peptide indicates that SmSPO-1 is a secreted protein or a membranal protein. In GenBank we found a schistosome cDNA sequence (accession number AFO91509) de-

Fig. 3. SmSPO-1 cDNA and gene. (Top) The cDNA sequence is a composite of 494 bp of the SmSPO-1 cDNA clone (T at position −12 to end of the poly(A) tail) and the upstream 17 bp to the cap site A at position −29 determined by primer extension sequencing. The coding region starts with the initiator ATG codon (A number +1). Three asterisks indicate termination codons: two (TGA) precede the coding region and are in-frame with the initiator ATG, one (TAG) is at the end of the coding region. The polyadenylation signal (AATAAA) and the poly(A) addition site (A base position 469) are underlined. The 22-mer oligonucleotide (complementary to the sequence marked by a wavy line) used for primer extention, the position of introns (arrowheads), and the N-terminus of the presumed mature protein (Leu[19], dot) are indicated. The one letter code for amino acids is used. (Bottom) Scheme of the SmSPO-1 gene (3856 bp sequenced). Exons (open boxes), introns (I1, I2, I3) and their size (bp), the initiator Met codon (M1), the C-terminal Ser codon (S117), the cap (CAP) site, the poly(A) addition site (pAad), *Eco*RI and *Hae*III restriction sites are indicated.

posited by Valle et al., which is very similar to SmSPO-1. The AFO91509 and SmSPO-1 cDNAs show 100% identity at the amino acid level, and they differ in six bases: two at the 5′ untranslated region and four at the 3′ untranslated region. In addition, 16 bases at the 5′ end of SmSPO-1 are missing from the AFO91509 sequence. The amino acid sequence of the mature SmSPO-1 (99 residues) showed limited homology with segments of other proteins, e.g., 31% identity with bacterial hydrogenase over a stretch of 77 residues (protein of 371 residues [12]), 25% identity over a stretch of 97 residues with members of the stathmin family (proteins of about 180 residues [13]). The significance of these homologies is not yet clear.

Genomic DNA of *S. mansoni* was digested with several restriction enzymes and hybridized with the SmSPO-1 cDNA. The Southern blot showed a simple pattern of hybridizing bands that conformed with one gene copy per haploid genome (data not shown). The sizes of the hybridizing bands in the *Eco*RI and *Hae*III digests were in complete agreement with *Eco*RI and *Hae*III fragments of the cloned gene.

A genomic library of cercarial DNA (partial *Sau*3A digest) constructed in λEMBL4 phage was screened with the SmSPO-1 cDNA. Four independent clones were isolated and characterized by restriction enzymes. The clones had overlapping fragments. Hybridization with the SmSPO-1 cDNA showed that in all clones the entire cDNA was located on two *Eco*RI fragments. One clone that was further studied (3856 bp sequenced) showed that the cDNA from the cap site till the poly(A) addition site was spread over 2.1 kb genomic DNA, and it was interrupted by three introns of 1404 bp, 48 bp and 34 bp (Fig. 3). Introns were bounded by typical donor and acceptor splice signals. The four exons and the cDNA showed 100% sequence identity. A 1 kb segment 5′ to the initiator Met codon contained a typical TATA box (26 bases upstream from the cap site) and other promoter elements.

The authors thank Dr. R. Tarrab-Hazdai for sup-

ply of rabbit antiserum to sonicate of cercarial tegument. This work was supported by a research grant from the Muriel Fusfeld Research Award.

References

[1] J. Cherfas, New hope for vaccine against schistosomiasis, Science 251 (1991) 630–631.
[2] Z. Grossman, D. Ram, A. Markovics, R. Tarrab- Hazdai, F. Lantner, E. Ziv, I. Schechter, *Schistosoma mansoni*: stage specific expression of muscle specific genes, Exp. Parasitol. 70 (1990) 62–71.
[3] S. Neumann, E. Ziv, F. Lantner, I. Schechter, Regulation of hsp70 gene expression during the life cycle of the parasitic helminth *Schistosoma mansoni*, Eur. J. Biochem. 212 (1993) 589–596.
[4] D. Ram, Z. Grossman, A. Markovics, A. Avivi, E. Ziv, F. Lantner, I. Schechter, Rapid changes in the expression of a gene encoding a calcium-binding protein in *Schistosoma mansoni*, Mol. Biochem. Parasitol. 34 (1989) 167–176.
[5] R. Parvari, E. Ziv, F. Lantner, S. Tel-Or, Y. Burstein, I. Schechter, Analysis of chicken immunoglobulin light chain cDNA clones indicate a few germline Vλ genes and allotypes of the Cλ locus, EMBO J. 6 (1987) 97–102.
[6] F. Sanger, A.R. Coulson, B.G. Barrel, A.J.H. Smith, B.A. Roe, Single-stranded bacteriophage as an aid to rapid DNA sequencing, J. Mol. Biol. 143 (1980) 161–178.
[7] J. Corden, B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, P. Chambon, Promoter sequences of eukaryotic protein-coding genes, Science 209 (1980) 1406–1414.
[8] J.V. Maizel, Polyacrylamide gel electrophoresis of viral proteins, Methods Virol. 5 (1972) 179–246.
[9] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132.
[10] I. Schechter, D.J. McKean, R. Guyer, W. Terry, Partial amino acid sequence of the precursor of immunoglobulin light chain programed by mRNA in vitro, Science 188 (1975) 160–162.
[11] H. Nielsen, J. Engelbrecht, S. Brunak, G. von Heijne, Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites, Protein Eng. 10 (1997) 1–6.
[12] H.P. Klenk, R.A. Clayton, J.F. Tomb et al., The complete genome sequence of the hyperthermophilic sulphate-reducing archaeon *Archaeoglobus fulgidus*, Nature 390 (1997) 364–370.
[13] S. Ozon, A. Maucuer, A. Sobel, The stathmin family, molecular and biological characterization of novel mammalian proteins expressed in the nervous system, Eur. J. Biochem. 248 (1997) 794–806.